University of Groningen

Porphyromonas gingivalis, the beast with two heads

Gabarrini, Giorgio

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Gabarrini, G. (2018). *Porphyromonas gingivalis, the beast with two heads: A bacterial role in the etiology of rheumatoid arthritis*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 04-05-2023

Chapter 3

Porphyromonas gingivalis – the venomous bite of an oral pathogen

Giorgio Gabarrini, Stefano Grasso, Arie Jan van Winkelhoff, and Jan Maarten van Dijl

Under consideration in *Microbiology and Molecular Biology Reviews*.

Abstract

Porphyromonas gingivalis is a renowned oral pathogen responsible for the extensively widespread disease periodontitis. In recent years, however, this bacterium has been implicated in the etiology of another common disorder, the autoimmune disease rheumatoid arthritis. Periodontitis and rheumatoid arthritis were known to correlate for decades, but only recently the lynchpin behind this association has been unveiled. *P. gingivalis* possesses an enzyme that citrullinates certain host proteins and, potentially, elicits autoimmune antibodies against such citrullinated proteins. These autoantibodies are highly specific for rheumatoid arthritis and have been purported not only as a symptom, but also as a potential cause of the disease. The citrullinating enzyme, and several other virulence factors of *P. gingivalis*, are targeted to the host tissue, either as secreted or outer membrane-bound proteins. As these virulence factors allow *P. gingivalis* to take a venomous bite out of the periodontium, the overall protein sorting and secretion events in this pathogen are of prime relevance for understanding its full disease-causing potential and for developing new preventive and therapeutic approaches. The aim of this review is therefore to offer a broad overview of the subcellular and extracellular localizations of all proteins in three reference strains and four clinical isolates of *P. gingivalis*, as well as the mechanisms employed to reach these destinations.

Introduction

Porphyromonas gingivalis is a Gram-negative, black-pigmented anaerobic bacterium belonging to the *Bacteroidetes* phylum. Although this bacterium is often described as rod-shaped, its appearance is more reminiscent of a little sausage (Fig. 1). *P. gingivalis* initially garnered interest for its role as a model organism for bacteria of the *Cytophaga-Flavobacterium-Bacteroides* (CFB) group and later on as an oral pathogen. This bacterium, is one of the major etiological agents of the oral disease periodontitis[1, 2], being present in almost 85% of severe cases[3]. Periodontitis is an inflammatory disorder affecting the tissue surrounding the teeth, the periodontium, potentially leading to tooth loss. Remarkably, the prevalence of periodontitis in the world population generally ranges from 10 to 20%, but this disease has been known to affect between 10 to 57% of different populations worldwide, depending on the degree of severity, socio-economic status and oral hygiene[4]. This extremely high incidence establishes periodontitis as one of the most common diseases in the world and as main cause of tooth loss worldwide[5, 6]. Additionally, periodontitis has been presented as a risk factor for, or associated to, several health conditions such as diabetes[7-12], heart diseases[13-15], dementia[16, 17], Alzheimer's disease[16-19], and especially rheumatoid arthritis (RA)[20-38].

RA is an inflammatory autoimmune disorder whose etiology is still not fully comprehended and that has been found to be clinically associated with periodontitis. In several countries, the prevalence of periodontitis appeared to be increased among RA patients in comparison with the general population[20, 25, 32, 33, 36, 39, 40]. Correspondingly, RA is more prevalent among patients with periodontitis[31-33, 36, 40], cementing the hypothesis of an intimate connection between the two disorders.

The suspected role of *P. gingivalis* in the interplay between periodontitis and rheumatoid arthritis has drawn attention to the bacterium's citrullinating enzyme[21, 23, 24, 28, 30, 37]. This enzyme, a peptidylarginine deiminase (PAD), catalyzes the conversion of arginine into citrulline residues in a post-translational protein modification called citrullination. Citrullination has the potential to alter the net charge of a substrate protein, possibly leading to severe changes in its structure and function[23].

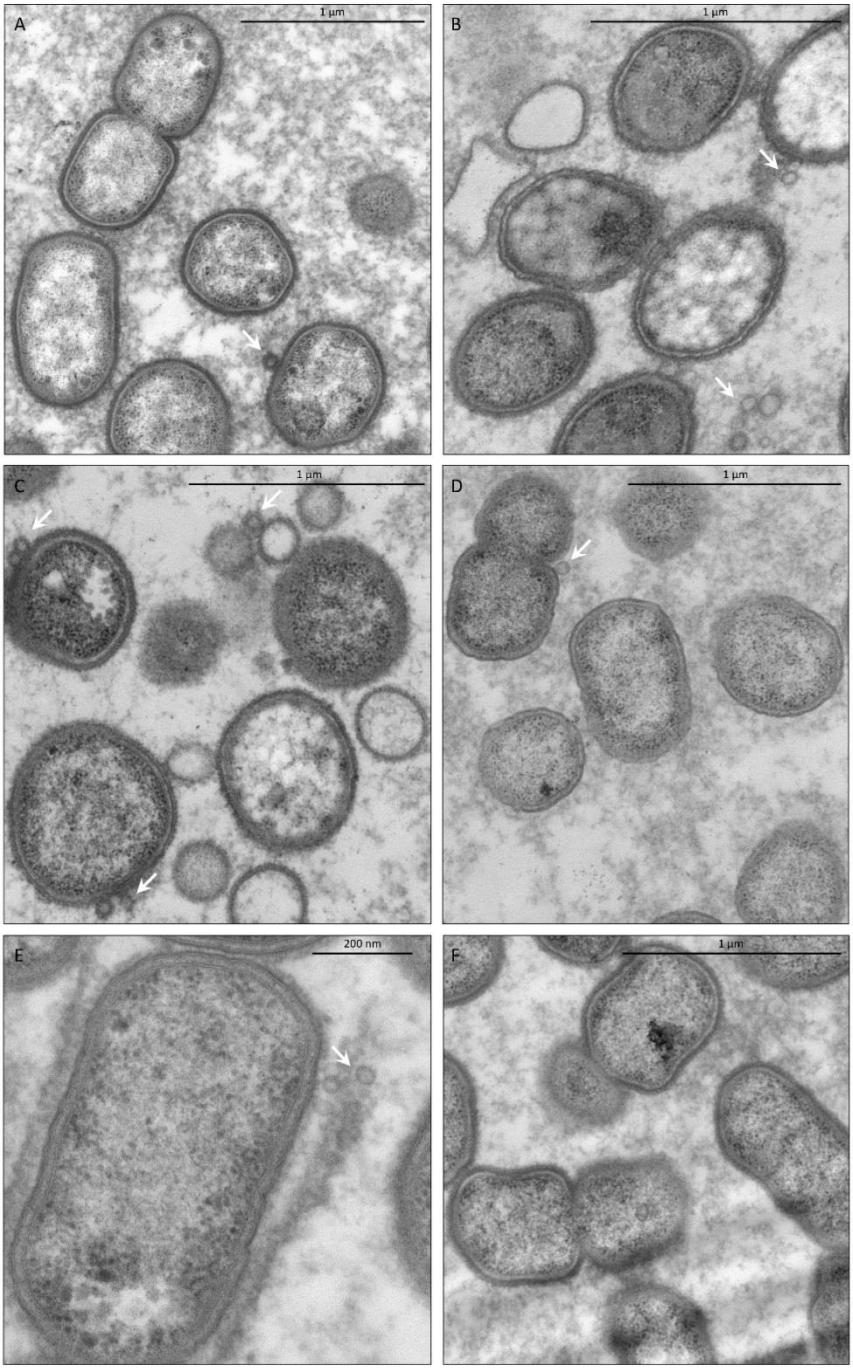

Figure 1. *Porphyromonas gingivalis*. Electron micrographs of *P. gingivalis* (A) type strain W83, and the clinical strains (B) 505700, (C) 512915, (D) 505759, and

(E-F) MDS33. Note the capture of OMV formation in panels A, B, C and E as marked by white arrows.

Although citrullination is a physiological process that takes place in a wide variety of healthy tissues as a general regulatory mechanism, especially during apoptosis, it also occurs in association with inflammatory processes.

While peptidylarginine deiminases are highly conserved in mammals, only three bacteria of the genus *Porphyromonas* are known to produce such enzymes[23, 34, 41-43]. The PAD of *P. gingivalis* (PPAD) and the homologous enzymes from *Porphyromonas loveana* and *Porphyromonas gulae* share no evolutionary relationship with the mammalian PADs[43, 44]. Remarkably, PPAD is believed to citrullinate certain human host proteins that, especially in genetically predisposed subjects[23, 30], can stimulate the production of anti-citrullinated protein antibodies (ACPAs)[23, 27, 34, 35, 41]. These ACPAs are 95% specific and 68% sensitive for RA[45, 46]. Interestingly, like many other bacterial virulence factors, the citrullinating enzyme of *P. gingivalis* is targeted to the host milieu, where it is detectable both in a secreted soluble state and in an outer membrane vesicle (OMV)-bound state[41, 47, 48]. In addition, a substantial portion of PPAD remains associated with the bacterial cell in an outer membrane (OM)-bound state. Other proteins that play a role in *P. gingivalis* colonization of the periodontal pockets, such as gingipains, hemagglutinins and fimbrial components, are exposed on the bacterial cell surface[49]. These findings and the possible role of *P. gingivalis* in the etiopathogenesis of RA focus interest on the mechanisms and pathways responsible for protein export in this oral pathogen.

Knowledge of the subcellular localization of proteins is an invaluable tool for genome annotation and the interpretation of proteomics data. The presence of a protein in a specific subcellular compartment can, in fact, hint at its function. Secreted proteins, for example, are expected to be involved in processes that require activities at the cell surface or beyond, such as nutrient acquisition, cell motility, cell-cell communication or host colonization and invasion. In particular, the surface proteins can represent sterling candidates for drug targets. This review will therefore provide a broad overview on protein localization in *P. gingivalis*, which is based on in-depth bioinformatic analyses of previously published biochemical, genomic and proteomic studies[47, 50-76].

General architecture and subcellular compartments

To predict the subcellular localization of proteins in a bacterium, it is paramount to first gather information on this bacterium's cellular architecture. The knowledge of subcellular compartments is in fact fundamental to develop a species-specific prediction strategy. In this review, a total of seven *P. gingivalis* strains was evaluated: three reference strains and four clinical isolates. The three investigated reference strains W83, TDC60 and ATCC33277 are the main and best-studied *P. gingivalis* strains with publicly available genome sequences that were manually curated. Their proteomes were accessed and downloaded from UniprotKB[77] on 11th March 2018: W83 [UP000000588], ATCC 33277 [UP000008842], and TDC60 [UP000009221]. The included clinical isolates (20655, MDS140, MDS33, 512915)[42, 48] can be divided in PPAD sorting types I and type II[48]. This classification concerns the differential sorting of PPAD as recently detected among clinical isolates. Compared to sorting type I isolates, sorting type II isolates display an extremely hampered production of OM- and OMV-bound PPAD, allegedly due to an amino acid substitution at position 373[48]. The two sorting type I isolates, 20655 and MDS140, were obtained from a patient with severe periodontitis but no RA, and a healthy carrier, respectively[42]. The sorting type II isolates (512915 and MDS33), on the other hand, were isolated from a periodontitis patient without RA and a patient with severe periodontitis and RA, respectively[48].

Considering its status as a Gram-negative bacterium, the protein-containing subcellular compartments of a *P. gingivalis* cell can be divided in cytoplasm, inner membrane (IM), periplasm, and OM (Fig. 2). Nascent bacteriophages could, in principle, represent a separate intracellular compartment, but to date no bacteriophages have been described for *P. gingivalis*[78, 79]. In addition, some of the proteins are targeted to the extracellular milieu, in particular the periodontium of the human host. As mentioned above, proteins can be secreted either in a soluble state or bound to OMVs (Fig. 2)[43, 48, 80, 81]. Gram-negative bacteria produce OMVs by natural "blebbings" of their outer membrane. Accordingly, OMVs consist of a single membrane originating from the OM, contain OM proteins, lipopolysaccharide (LPS) and other lipids. The cargo of OMVs usually encompasses cytoplasmic and periplasmic proteins[82], but they appear enriched in

virulence factors[80]. The OMVs of *P. gingivalis* seem to accumulate in gingival tissue at diseased sites in chronic periodontitis patients, but not at healthy sites[83]. Notably, the OMVs should be regarded as important virulence factors since they can enter host epithelial cells and degrade key receptor proteins using specific cysteine proteinases called gingipains[84-86]. Gingipains are essential for virulence in animal models where they were shown to degrade many host proteins, thereby impairing cellular functions and the host immune response[87, 88]. Therefore, OMVs represent 'undead' satellite compartments of the *P. gingivalis* cell (Figs. 2 and 3).

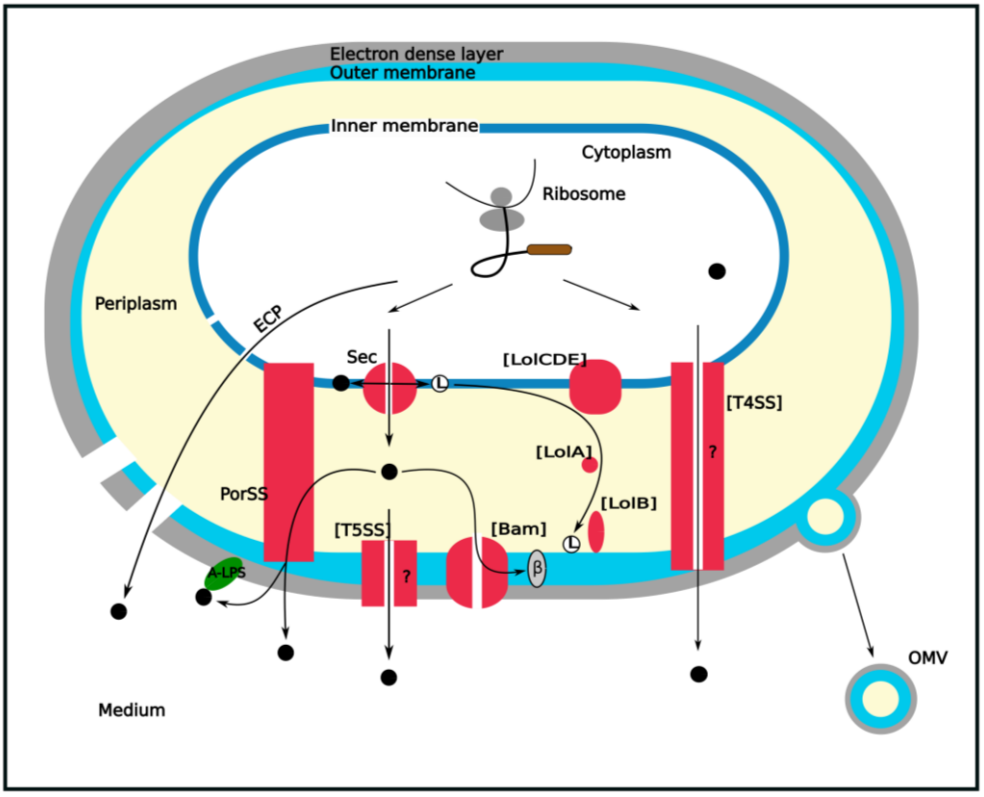

Figure 2. *Sorting mechanisms in P. gingivalis*. Overview of the cell architecture and the transport systems occurring in *P. gingivalis*, according to domain analyses of major components of known transport systems in Gram-negative bacteria.

Of note, however, the precise mechanism underlying the formation of OMVs in *P. gingivalis* is still unknown and therefore no bioinformatic tool exists or can be created yet to find the proteins located in such a peculiar extracellular compartment. While

biochemical studies have investigated the OMV cargo proteins[80], their results are limited to the one strain analyzed. For this reason, the OMV compartment has not been taken into account in our bioinformatic appraisal of the available data.

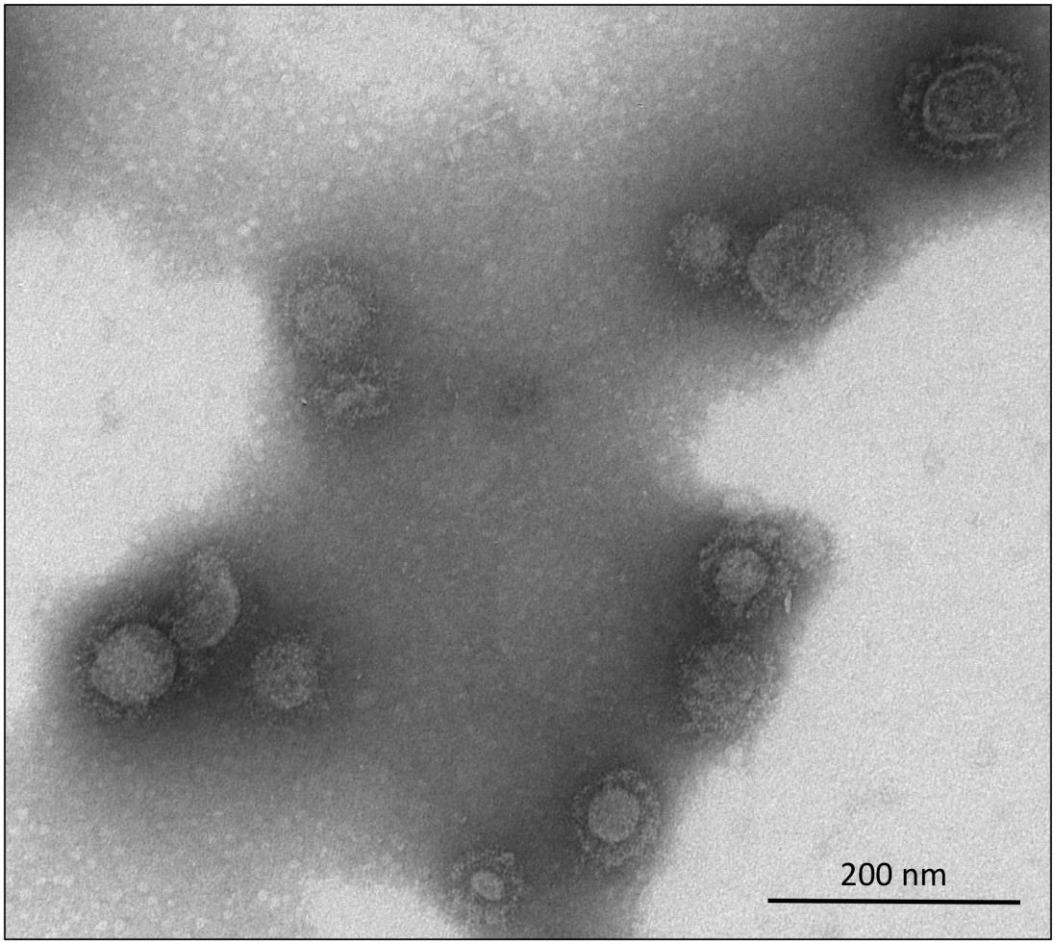

Figure 3. *OMVs are a satellite compartment of P. gingivalis*. Electron micrograph of purified outer membrane vesicles of *P. gingivalis* type strain W83.

Systems for protein export from the cytoplasm, membrane insertion and secretion in *P. gingivalis*

Knowledge of the subcellular compartments present in *P. gingivalis* is required for the identification of transport, secretion and

membrane insertion systems. In general, Gram-negative bacteria possess an IM and an OM and, for this reason, several export and membrane insertion systems are utilized to translocate proteins across these two membranes or to sort them to their rightful destination. The vast majority of extracytoplasmic proteins in Gram-negative bacteria are translocated across the cytoplasmic membrane in an unfolded state by the Sec translocase, including inner membrane lipoproteins[89] (Fig. 2). A smaller number of proteins traverses the cytoplasmic membrane *via* the Tat system, which is specific for cargo proteins in a pre-folded state[90-92]. Among the proteins that have reached the periplasm, β-barrel proteins can then be inserted into the outer membrane by the 'β-barrel assembly machinery' (BAM) complex[93, 94], while lipoproteins are inserted by the 'localization of lipoprotein' (Lol) system[95, 96] (Fig. 2). Due to their major role in protein sorting, these systems are broadly conserved among Gram-negative bacteria and their genes are, thus, easily recognizable by automated pipelines. Key components of the Sec, Tat, BAM and Lol systems can be promptly identified by looking for the homologues of known members of these systems in other species. In addition, specific domain searches can be utilized (Table 1).

With the exceptions of SecE and SecG, either missing among some clinical strains or poorly annotated, all the analyzed strains of *P. gingivalis* possess the components of the SecYEG-DFyajC system. Intriguingly, however, they lack the known Tat translocase components, showing that the Tat system is absent in this bacterium. This is consistent with the outcome of domain searches using motifs identifying Tat signals (Tigr01409, Tigr01412, pfam10518), which yield no matches (Table 1), and with previous analyses reported in the literature[52]. Moreover, only two members of the BAM system (BamA and BamC) appear to be present in the strains studied, as shown by domain searches and similarity analyses (Table 1). No homologues of BamB, BamD and BamE are present in *P. gingivalis*. This is noteworthy, because only BamA and BamD are regarded universally essential for functionality of the Bam system[97]. Similarly, the Lol system is only partially represented in *P. gingivalis* as there are merely four proteins with a LolE motif (COG4591) in the *P. gingivalis* reference strains. Some of these proteins display moderate levels of similarity to LolE proteins from other Gram-negative species, including potential LolCE motifs (tigr002212, tigr002213).

Table 1. *Presence or absence of known protein transport and membrane insertion systems in P. gingivalis.* Key members of protein transport systems and membrane protein insertion systems in the three *P. gingivalis* reference strains were identified by domain searches and secondary verification of the presence of particular orthologs.

SS	DOMAIN	PROTEIN	ATCC 33277	W83	TDC60
Sec	COG0653	SecA	PGN_1458	PG0514	PGTDC60_1633
	Tigr00963	SecA	PGN_1458	PG0514	PGTDC60_1633
	IPR003708	SecB			PGTDC60_1688
	IPR035958	SecB			
	IPR027398	SecD first TM region			
	IPR005791	SecD	PGN_1702	PG1762	PGTDC60_1374
	IPR005665	SecF	PGN_1702	PG1762	PGTDC60_1374
	IPR022645	SecD/F			
	COG0690	SecE	PGN_1577	PRESENT*	PGTDC60_1503
	Tigr00964	SecE	PGN_1577	PRESENT*	PGTDC60_1503
	COG1314	SecG	PGN_0258	PG0144	PGTDC60_0422
	Tigr00810	SecG	PGN_0258	PG0144	PGTDC60_0422
	COG0201	SecY	PGN_1848	PG1918	PGTDC60_0188
	Tigr00967	SecY	PGN_1848	PG1918	PGTDC60_0188
	COG0706	YidC	PGN_1446	PG0526	PGTDC60_1645
	Tigr03592	YidC	PGN_1446	PG0526	PGTDC60_1645
	Tigr03593	YidC	PGN_1446	PG0526	PGTDC60_1645
	COG1862	YajC	PGN_1485	PG0485	PGTDC60_1601
	Tigr00739	YajC	PGN_1485	PG0485	PGTDC60_1601
Sec			+	+	+
TAT	COG0805	TatC			
	Tigr00945	TatC			
	pfam00902	TatC			
	COG1826	TatA/E			
	Tigr01411	TatA/E			
	Tigr01410	TatB			
	Tigr01409	Tat signal			
	Tigr01412	Tat signal			
	pfam10518	Tat signal			
TAT			-	-	-
SRP	IPR004780	Ffh	PGN_1205	PG1115	PGTDC60_1100
	IPR004390	FtsY	PGN_0264	PG0151	PGTDC60_0428
SRP			+	+	+
BAM	Tigr03303	BamA	PGN_0299	PG0191	PGTDC60_0462
	IPR023707	BamA			

	Tigr03300	BamB			
	IPR017687	BamB			
	IPR014524	BamC			
	Tigr03302	BamD	PGN_1354	PG1215	PGTDC60_1188
	IPR017689	BamD	PGN_1354	PG1215	PGTDC60_1188
	pfam06804	BamD			
	pfam04355	BamE			
	IPR026592	BamE			
BAM			±	±	±
LOL	Tigr00547	LolA			
	pfam03548	LolA			
	COG2834	LolA			
	Tigr00548	LolB			
	COG3017	LolB			
	pfam03550	LolB			
	Tigr02212	LolC			
	Tigr02211	LolD			
	Tigr02213	LolE			
	COG4591	LolE	PGN_0718	PG0682	PGTDC60_0845
			PGN_0719	PG0683	PGTDC60_1224
			PGN_1025	PG0922	PGTDC60_1807
			PGN_1387	PG1252	PGTDC60_1808
LOL			±	±	±
T1SS	Tigr01842	PrtD			
	IPR010128				
	Tigr01843	HlyD			
	IPR010129				
	Tigr01844	TolC			
	IPR010130				
	Tigr01846	HlyB			
	IPR010132				
	Tigr03375	LssB			
	IPR017750				
	pfam02321	outer membrane efflux protein	PGN_0444	PG0063	PGTDC60_0345
			PGN_0715	PG0094	PGTDC60_0374
			PGN_1432	PG0285	PGTDC60_0631
			PGN_1539	PG0538	PGTDC60_1397
	IPR003423		PGN_1679	PG0679	PGTDC60_1540
			PGN_2012	PG1667	PGTDC60_1656
			PGN_2041		PGTDC60_1804
	pfam03412	bacteriocin exporter family			PGTDC60_1000

	IPR005074	(Peptidase C39 family)			PGTDC60_1973
T1SS			-	-	-
T2SS	COG1450	PulD			
	COG2804	PulE			
	COG1459	PulF			
	COG2165	PulG			
	IPR013545	PulG			
	Tigr02517	type II secretion system protein D (GspD)			
	IPR013356				
T2bSS	Tigr02519	pilus (MSHA type) biogenesis protein MshL			
	IPR013358				
	Tigr02515	type IV pilus secretin (or competence protein) PilQ			
	IPR013355				
	pfam07655	Secretin N-terminal domain			
	IPR011514				
	pfam07660	Secretin and TonB N terminus short domain			
	IPR011662				
T2a-cSS, T3aSS	pfam00263	Bacterial type II and III secretion system protein (secretin)			
	IPR004846				
	pfam03958	Bacterial type II/III secretion system short domain			
	IPR005644				
T2SS			-	-	-
T3SS	COG1157	FliI			
	IPR032463	FliI			
	COG1766	FliF			
	IPR000067	FliF			
	COG1886	FliN			
	IPR012826	FliN			
T2a-cSS, T3aSS	pfam00263	Bacterial type II and III secretion system protein (secretin)			
T2a-bSS, T3aSS	pfam03958	Bacterial type II/III secretion system short domain			
T3aSS	Tigr02516	type III secretion outer membrane pore, YscC/HrcC family			
	IPR003522				
T3bSS	pfam02107	Flagellar L-ring protein (FlgH)			
	IPR000527				
T3SS			-	-	-
T4SS	COG3838	VirB2			
	IPR007039	VirB2			
	COG3702	VirB3			
	IPR007792	VirB3			

	COG3451	VirB4	PGN_0065	PG1481	PGTDC60_1018
					PGTDC60_1993
	COG3704	VirB6			
	IPR007688	VirB6			
	COG3736	VirB8			
	IPR007430	VirB8	PGN_0062	PRESENT*	PGTDC60_1021
	IPR026264	Type IV secretion system protein VirB8/PtlE			
	COG3504	VirB9			
	IPR014148	VirB9			
	COG2948	VirB10			
	IPR005498	VirB10			
	COG0630	VirB11			
	IPR014155	VirB11			
	COG3505	VirD4	PGN_0076	PG1490	PGTDC60_1006
			PGN_0579		PGTDC60_1984
	IPR003688	Type IV secretion system protein TraG/VirD4		PG1490	PGTDC60_1984
T4bSS	pfam03524	Conjugal transfer protein			
	IPR010258				
	Tigr02756	type-F conjugative transfer system secretin TraK			
	IPR014126				
	pfam06586	TraK protein			
	IPR010563				
T4SS			±	±	±
T5SS	COG3468	adhesin AidA			
	COG5295	autotransporter adhesin			
	COG5571	autotransporter β-barrel domain			
T5cSS	pfam03895	YadA-like C-terminal region			
	IPR005594				
T5aSS	pfam03797	autotransporter β domain			
	IPR005546	autotransporter β domain	PGN_0129	PG1823	PGTDC60_0070
			PGN_0178	PG2130	PGTDC60_1255
			PGN_1744	PG2168	PGTDC60_1292
T5dSS	pfam07244	Surface ag VNR domain (PlpD POTRA motif)	PGN_0299	PG0191	PGTDC60_0462
	IPR010827				
	pfam01103	Bacterial surface Ag domain (PlpD β-barrel domain)	PGN_0147	PG0980	PGTDC60_0900
	IPR000184		PGN_0973	PG2095	PGTDC60_1324
T5SS			-	-	-
T5dSS			±	±	±
T6SS	Tigr03345	type VI secretion ATPase, ClpV1 family			
	IPR017729				

	Tigr03347	type VI secretion protein, VC_A0111 family			
	IPR010732				
	Tigr03350	type VI secretion system OmpA/MotB family protein			
	IPR017733				
	Tigr03352	type VI secretion lipoprotein, VC_A0113 family			
	IPR017734				
	Tigr03353	type VI secretion protein, VC_A0114 family			
	IPR010263				
	Tigr03354	type VI secretion system FHA domain protein			
	IPR017735				
	Tigr03355	type VI secretion protein, EvpB/VC_A0108 family			
	IPR010269				
	Tigr03358	type VI secretion protein, VC_A0107 family			
	IPR008312				
	Tigr03362	type VI secretion-associated protein, VC_A0119 family			
	IPR017739				
	Tigr03373	type VI secretion-associated protein, BMA_A0400 family			
	IPR017748				
T6SS			-	-	-
T7SS	Tigr03919	type VII secretion protein EccB			
	IPR007795				
	Tigr03920	type VII secretion integral membrane protein EccD			
	IPR006707				
	Tigr03921	type VII secretion-associated serine protease mycosin			
	IPR023834				
	Tigr03922	type VII secretion AAA-ATPase EccA			
	IPR023835				
	Tigr03923	type VII secretion protein EccE			
	IPR021368				
	Tigr03924	type VII secretion protein EccCa			
	IPR023836				
	Tigr03925	type VII secretion protein EccCb			
	IPR023837				
	Tigr03926	type VII secretion protein EssB			
	IPR018778				
	Tigr03927	type VII secretion protein EssA/YueC			
	IPR018920				
	pfam10661	WXG100 protein secretion system (Wss), EssA			
	IPR034026				
	Tigr03928	type VII secretion protein EssC			
	IPR023839				

	IPR022206				
	Tigr03931	type VII secretion-associated protein, Rv3446c family			
	IPR023840				
	pfam00577	Fimbrial usher protein			
	IPR000015				
	pfam06013	Proteins of 100 residues with WXG			
	IPR010310				
T7SS			-	-	-
T8SS (ENP)	pfam03783	Curli production assembly/transport component CsgG			
	IPR005534				
	pfam07012	Curlin associated repeat			
	IPR009742				
	pfam10614	Tafi-CsgF			
	IPR018893				
	pfam10627	CsgE			
	IPR018900				
T8SS (ENP)			-	-	-
PorSS		PorK	PGN_1676	PG0288	PGTDC60_1400
		PorL	PGN_1675	PG0289	PGTDC60_1401
		PorM	PGN_1674	PG0290	PGTDC60_1402
		PorN	PGN_1673	PG0291	PGTDC60_1403
		PorP	PGN_1677	PG0287	PGTDC60_1399
		PorQ	PGN_0645	PG0602	PGTDC60_1728
	pfam13568	PorT	PGN_0778	PG0751	PGTDC60_1868
		PorU	PGN_0022	PG0026	PGTDC60_0023
		PorV (PG27,LptO)	PGN_0023	PG0027	PGTDC60_0024
		PorW	PGN_1877	PG1947	PGTDC60_0218
	pfam14349	Sov	PGN_0832	PG0809	PGTDC60_1927
		PorX	PGN_1019	PG0928	PGTDC60_0851
		PorY	PGN_2001	PG0052	PGTDC60_0334
		Lipoprotein; TPRd, WD40d, CRDd, OmpA family domain	PGN_1296	PG1058	PGTDC60_0980
		PorZ	PGN_0906	PG1064	PGTDC60_1144
	Orthology	PorZ	PGN_0509	PG1604	PGTDC60_0697
		b-barrel protein	PGN_0297	PG0189	PGTDC60_0460
		TonB-dependent receptor; b-barrel protein	PGN_1437	PG0534	PGTDC60_1652
		Omp17; OmpH-like	PGN_0300	PG0192	PGTDC60_0463
		sigP	PGN_0274	PG0162	PGTDC60_0438
PorSS			+	+	+

These proteins are predicted as inner membrane proteins, which would be consistent with the localization of the LolCE proteins of *Escherichia coli,* and half of them belong to the core proteome. Yet, canonical members of the Lol system, especially LolA, LolB, LolC and LolD are absent from *P. gingivalis*. These observations suggest that analogous 'Lol' and 'BAM' systems may, respectively, be operational in the IM and OM of this bacterium (Fig. 2), while the prototype Lol and BAM systems are lacking.

Gram-negative bacteria can also possess other common systems enabling the translocation of proteins across the OM[98]. These secretion systems vary from type I to type VIII (T1SS-T8SS), with the addition of a type IX specific to certain members of the *Bacteroidetes* phylum[52]. The type IX system is also referred to as T9SS, or Porin secretion system (PorSS), and the latter designation is most frequently used in the context of protein export in *P. gingivalis*. Unfortunately, secretion systems are not usually well annotated by automated pipelines, mainly because certain members of different secretion systems (e.g T2SS and T4SS) share a higher sequence similarity with one another than functionally equivalent members of the same secretion system (e.g. pilin proteins). Moreover, many secretion systems are still poorly characterized, leading to difficulties in finding the most suited domains for a domain search. Fortunately, the genes encoding members of these systems usually co-localize on the genome, thus facilitating the identification of system components.

The potential presence of known secretion systems in *P. gingivalis* was revaluated *via* domain searches, literature and genome context analyses and similarity searches across the *P. gingivalis* reference strains. All three analyzed reference strains lack the vast majority of secretion systems commonly encountered in Gram-negative bacteria (Table 1). Nevertheless, proteins containing two motifs belonging to members of the type I secretion system, pfam02321 and pfam03412, were found. The pfam02321 motif was detected in multiple proteins across all strains while pfam03412 was present only in two proteins for the TDC60 strain. Interestingly, these two proteins display no significant similarity to proteins in the other *P. gingivalis* reference strains as opposed to a significant similarity shared with proteins belonging to other species in the same phylum. Of note, the pfam02321 motif can also detect OM components of drug and metal efflux pumps, suggesting that the identified proteins do not belong to a functional type I secretion system. Conversely, the pfam03412

motif was used in combination with tigr01193 to identify bacteriocin exporters and the two proteins identified in strain TDC60 appear to possess both, although the scores for tigr01193 are significantly lower than the one for pfam03412. In conclusion, it thus appears that a type I secretion system is absent from *P. gingivalis*[52].

None of the known protein components of type II and III secretion systems was found in *P. gingivalis*, including members of subclasses a, b, and c of type II and subclasses a and b of type III secretion systems. Conversely, domain searches of three major components of the type IV secretion system, VirB4, VirB8, and VirD4, showed multiple matches across the three reference strains. The VirB4 domain is present in two TDC60 proteins sharing a relatively high level of similarity, while the VirD4 domain is present in two W83 and two TDC60 proteins. At least one gene per strain encoding these proteins colocalizes with a VirB4 motif gene on the *P. gingivalis* chromosome with a distance of about 5 kb for the respective W83 genes, 7 kb for the ATCC 33277 genes and 10-11 kb for the TDC60 genes. The VirB8 domain is present in one gene per reference strain, albeit the annotation used did not reveal a gene for the W83 strain and the presence of the domain was discovered after genome analysis. Similarly, albeit no matches were found for signature domains of key components of the type V secretion system, one protein in every reference strain was found to display a PlpD motif (pfam07244). This motif identifies components of subclass d of the type V secretion system. Moreover, these proteins appear to possess a second PlpD motif used in T5dSS searches, pfam01103, although with a sensibly lower score. Altogether, it is difficult to predict the activity of T4SS and T5dSS in the analyzed *P. gingivalis* strains. In the canonical T4SS, VirB4 and VirD4 are two of the three ATPases energizing the secretion machinery[99], which shapes the hypothesis that they may be used by another secretion system or for another function. This renders the possibility that a T4SS could function in *P. gingivalis* in absence of the other key members less likely. Another potential piece of evidence reinforcing this view is the fact that the VirB8 domain, used for this analysis, also recognizes conjugal transfer proteins like TrbF and TraK. However, VirB8 is generally responsible for forming the channel thanks to which the T4SS cargo proteins are translocated through the inner membrane. Hence the detected proteins containing a VirB8 domain could potentially offer an alternative pathway to the Sec system for protein passage across the IM of *P. gingivalis* (Fig. 2).

No proteins belonging to secretion systems VI, VII and VIII were detectable in the *P. gingivalis* strains analyzed, which is in agreement with previous literature[52]. On the other hand, *P. gingivalis* strongly relies on a novel secretion system shared by members of the *Bacteroidetes* phylum, namely the afore-mentioned PorSS [52] (Fig. 2). The PorSS comprises several proteins broadly conserved throughout the *Bacteroidetes* group and is also involved in gliding motility in many species of this phylum. As of now, there is general consensus that 17 proteins are essential for the PorSS function, although two additional proteins are deemed essential as well [75, 76]. Four of these proteins form the PorSS core membrane complex PorK-N. PorM, the main component of this complex, appears to localize in the IM, together with PorL. However, PorM is capable of interacting with the rest of the complex comprising the OM-bound lipoprotein PorK and the periplasmic OM-bound protein PorN thanks to its long periplasmic domain[52, 100]. Proteins that are exported *via* the PorSS are targeted for secretion by conserved C-terminal domains (CTDs)[52], which can be identified by the TIGR04131 (IPR026341) and TIGR04183 (IPR026444) motifs. The presence of a Sec-type N-terminal signal peptide in proteins exported by PorSS suggests that these proteins are exported across the IM by the Sec machinery. Importantly, all known members of the Por system are present in all of the strains evaluated here, highlighting the major role that this export system plays in *P. gingivalis* (Fig. 2).

It is clear that certain CTD proteins cross the outer membrane with the help of the PorSS, subsequently appearing both in the OM and in the secreted fractions as shown by different localization studies[47, 101, 102]. According to the current models, the CTD is cleaved during the export of the CTD proteins by a sortase-like mechanism and the resulting proteins can be secreted or re-attached to the OM *via* A-LPS modification, with the A-LPS acting as an anchor to the bacterial surface[75, 76]. Consistent with this model, the CTD is lacking from the mature soluble forms of these proteins[103, 104], and it has not been detected in OM-associated forms, which are extensively A-LPS modified[105-107]. Clearly, especially due to the two possible destinations of CTD proteins (i.e. OM insertion or secretion), it is challenging to predict the precise localization of CTD proteins by bioinformatic approaches.

In addition to the classical secretion systems, the creation and release of the aforementioned OMVs (Fig. 3) should be regarded as a specialized secretion pathway dedicated to virulence and the capture

of nutrients[80]. Indeed, the mechanism triggering the blebbing process that leads to these nanostructures, albeit poorly understood, is not random[108]. Additionally, the proteins secreted *via* this pathway seem to empanel mostly periplasmic and OM proteins, which serve as virulence factors[80]. Among the latter, proteases, especially the gingipains, appear to be most abundant[80]. The prevalence of proteases in the OMVs might serve several purposes. Firstly, it could be a way to deliver them to their foreign targets, especially proteins of phagocytic cells. Secondly, encasing the proteases within the membrane of the OMVs can protect them and/or the rest of the OMVs cargo from the outside environment, either physically or by rendering proteolytic sites on these proteins inaccessible. Thirdly, this feature might have evolved to protect *P. gingivalis* proteins, for example bound to the OM, from the bacterium's own highly proteolytic potential. Lastly, OMVs could serve a decoy function in immune evasion by *P. gingivalis*. In light of these different scenarios, the observed phenomenon of extracellular compartmentalization through vesiculation might be categorized as a 'protective secretion' behavior. In fact, the attachment of CTD proteins like PPAD to the OM and OMVs' membrane *via* A-LPS modification seems to protect them from proteolysis by *P. gingivalis*' own proteases, as evidenced by the recent observation that OMV disruption by ultrasound results in PPAD degradation[109]. In addition, the finding that PPAD sorting type II isolates, which are ineffectively attached to the OM and OMVs, are processed to a 37 kDa form is consistent with the idea that the OMVs serve to protect cargo against proteolysis[48].

Signal peptidases

Identification of the secretion system suite in *P. gingivalis* warranted a further investigation on the signal peptidases involved. Firstly, the Sec system utilizes two different types of signal peptidases. In general, cargo proteins of the Sec pathway are processed by signal peptidase I, which belongs to the S26 Merops family[110-112]. As is the case for all living cells, the *P. gingivalis* genome encodes signal peptidase I, as identified through COG0681 and pfam00717 searches. Of note, a recent study from Bochtler *et al.* showed that over 60% of signal peptidase I substrates in *P. gingivalis* display a glutamine residue immediately downstream of the signal peptidase I cleavage site (in position +1), irrespective of their subcellular localizations[113].

These glutamine residues are cyclized to pyroglutamate residues by the glutaminyl cyclase PG2157 (alternatively called PG_RS09565), a lipoprotein most likely located in the IM[113]. This high frequency of signal peptidase I substrates with a glutamine residue in position +1 is a common feature of most *Bacteroidetes* species[113].
Lipoproteins have N-terminal signal peptides recognized and removed by the signal peptidase II, which takes place after the invariant cysteine residue at position +1 relative to the cleavage site has been diacyl-glyceryl modified by the diacyl-glyceryl transferase Lgt[111]. Signal peptidase II belongs to the A08 Merops family and is detectable in *P. gingivalis* through domain searches for pfam01252 and COG0597 motifs. Likewise, Lgt is conserved in all investigated *P. gingivalis* strains as confirmed by BLAST searches. In *E. coli*, the N-terminal amino group of the diacyl-glyceryl-modified cysteine of the mature lipoprotein is acylated by the N-acyl transferase Lnt[111]. This may not be the case in *P. gingivalis*, as no homologues of the *E. coli* Lnt were detected in the investigated strains. However, the possibility of N-acylation of the mature lipoprotein upon cleavage by Lsp cannot be fully excluded, since it was shown that N-acylation by an as yet unidentified enzyme takes place in *Staphylococcus aureus*[114]. Interestingly, the N-acylation of staphylococcal lipoproteins has been invoked in the silencing of innate and adaptive immune responses[114], which is a trait that could enhance the fitness and pathogenicity of *P. gingivalis* as well.

Available algorithms for genome-wide identification of exported bacterial proteins

Genome-wide prediction of the subcellular localization of proteins is a relatively recent endeavor in proteomics, but not one without following[115-118]. Various bioinformatic tools have been designed to identify signal peptides, such as SignalP[119], Predisi[120] and Phobius[121]. These algorithms are generally used to predict signal peptides cleaved by signal peptidase I, but they do not readily recognize the lipoprotein signal peptides that are cleaved by signal peptidase II. To address this issue, lipoproteins have to be identified first by predictors capable of recognizing lipoprotein signals, such as LipoP[122] and Lipo[123]. The subsequently developed PSORT I represented the first comprehensive bacterial protein localization predictor. Since then, several prediction tools for protein localization have been

developed and implemented, rendering bioinformatic approaches a viable alternative to biochemical localization studies [118, 124, 125]. All these studies have in common the development of a complex network of subcellular localization predictors tailored on a specific bacterium in order to predict, as accurately as possible, the position of each protein in the proteome. One of such studies[125] has been taken into particular consideration for this review, and its workflow was adapted to review the overall protein localization in *P. gingivalis*.
All publicly available prediction tools for protein subcellular localization have particular pro's and con's. One of the difficulties in selecting the most suited programs for a bacterium of interest lies in the fact that publicly available predictors may quickly cease to be maintained, are subject to major modifications, or even become obsolete. This, coupled with the fact that certain programs may be more suited to bacteria of a certain group, makes it difficult to implement strategies previously developed for major model organisms, such as *E. coli* or *Bacillus subtilis*[111, 115, 117, 126]. Aside from public access, another important parameter determining our choice of programs was availability of a batch submission option, which grants fast genome-wide analyses. Moreover, to further refine the selection of prediction programs for a comprehensive overview of subcellular protein localization in *P. gingivalis*, tools with a high level of specialization were used as listed in Table 2.

Table 2. *List of localization predictors.* Overview of localization predictors, membrane insertion detectors, and other programs used in this study and their relative strengths and weaknesses.

NAME	USE	LIMITATIONS
LipoP	primarily prediction of Sec signal peptides that are cleaved by SpII but also provides prediction of inner membrane or cytoplasmic localization as well as SpI cleavage	does not detect Tat substrates
Lipo	prediction of Sec signal peptides cleaved by SpII	does not detect Tat substrates
SignalP	prediction of Sec signal peptides cleaved by SpI	does not detect Tat substrates
Predisi	prediction of Sec signal peptides cleaved by SpI	does not detect Tat substrates
Phobius	prediction of alpha helices in inner membrane proteins, distinguishing N-terminal TM from signal peptides	
TmHmm	prediction of alpha helices in inner membrane proteins	signal peptides often considered TM spans
Bomp	prediction of beta-barrel spans in outer membrane proteins	
SecretomeP	prediction of ECP	limited number of sequence per batch
Interpro	functional analysis of proteins by classification into families, domain and site prediction by combination of protein signatures from a number of member databases	

In most cases, such tools were single function predictors with few limitations, especially limitations that could have been offset by the application of other programs.
Intriguingly, different predictors occasionally assigned the same proteins to different subcellular compartments, even in case of programs with the same specific functions. Disagreements in localization between different programs underscore the notion that some predictors may be more accurate or, at the very least, better suited to chart the proteins of a specific bacterium than others. Moreover, these discrepancies reveal the levels of uncertainty of bioinformatics predictions and the need for an organized method encompassing all the chosen tools that can exploit all the strengths and balance the limitations of each program. On the other hand, one has to realize that protein sorting mechanisms in the living bacterium do usually not operate with a fidelity of 100%, which means that proteins that are generally secreted are detectable within different cellular compartments, while proteins that are meant to be retained in the cell (e.g. cytoplasmic proteins, lipoproteins or cell wall-bound proteins) can be encountered in the extracellular environment. The protein sorting ambiguities encountered *in silico* are thus perhaps an unintended reflection of the imperfections of sorting systems employed by a bacterial cell *in vivo*. Clearly, as long as these imperfections have no bearing on the competitive success of a bacterium, they do not matter.
Still, to meet the need for biologically relevant predictions of protein sorting, a decision tree (Fig. 4) was devised, organizing the predictors and sorting proteins through them with the purpose of assigning them to their rightful subcellular compartment. The first challenge in the prediction analysis is to localize the components of the export, secretion and membrane insertion systems themselves, which relates to the difficulty in recognizing their signal peptides by predictors. The level of difficulty depends on the system examined, with more common and conserved systems being more easily localized. Components of the Por secretion system, in fact, being only recently discovered, have in some cases an uncertain localization. Secondly, the identification of lipoproteins has priority, especially considering the inability of different predictors to distinguish Sec signal peptides cleaved by signal peptidase II, the lipoprotein-specific signal peptidase, from Sec signal peptides cleaved by signal peptidase I.

Figure 4. *Bioinformatics pipeline to unravel protein sorting events in P. gingivalis*. The flowchart depicts the different steps employed to assess the subcellular localization of proteins in the analyzed *P. gingivalis* strains.

Notably, localization tools generally distinguish between IM and OM lipoproteins utilizing data from extensive research done on the widely favored model Gram-negative bacterium *E. coli*. These studies have shown that lipoproteins possessing an aspartic acid in the position +2 of the mature protein become inner membrane lipoproteins (D^{+2} rule), while all the others are presented to the outer membrane by the Lol system[127]. Intriguingly, several exceptions to this rule have been found in other species[113, 126, 128-131] presenting the possibility that it is only obeyed by *Enterobacteriaceae*. Analyzing known outer membrane lipoproteins of *P. gingivalis* applying the D^{+2} rule, in fact, resulted in a faulty prediction of the subcellular location for the vast majority of lipoproteins. Conversely, alignment of lipoproteins of known subcellular localization showed a preferential glycine residue at position +2 or +3 of the mature form for inner membrane lipoproteins. The present evaluation of lipoprotein localization in *P. gingivalis* therefore relied on inspection of the +2 and +3 residues combined with specific domain searches. Following designation of the 'lipoproteome', investigation of proteins with transmembrane helices and Sec signal peptides was performed, in this order (Fig. 4). This relates also to the fact that predictors of membrane spanning regions occasionally mistake relatively longer signal peptides for transmembrane spans (Table 2).

Excretion of cytoplasmic proteins (ECP), also termed non-classical or leaderless secretion, is a highly discussed topic and a way to explain the presence in the extracellular milieu of proteins that lack a known signal peptide and a dedicated transport system of the categories described above[132, 133]. This quality applies to the bulk of cytoplasmic proteins. Accordingly, for a long time the most accredited hypothesis to explain the presence of such proteins in the extracellular milieu was cell lysis[134]. This view is supported by the observation that ECP can be associated with autolysin and phage activity, or the production of cytotoxic peptides[133]. Nonetheless, the existence of dedicated 'non-classical' secretion systems for proteins deprived of known signal peptides cannot be excluded, as underpinned by the relatively recent discovery of the Tat and type VII secretion systems[135]. Such hidden treasures are likely to be buried in the exoproteome haystack, until uncovered by the application of molecular biological or mass spectrometric approaches to assess bacterial protein secretion. In fact, with increasing sensitivity of mass spectrometric measurements, more and more signal peptide-less proteins have been identified in bacterial exoproteomes. This is exemplified by a recent investigation

on the exoproteome of *P. gingivalis*, where many signal peptide-less proteins were identified in the growth medium fraction[49]. In fact, this analysis highlights two remarkable features. Firstly, signal peptide-less extracellular proteins were overrepresented amongst the low-abundance extracellular proteins and the detection of these proteins was most variable between the investigated strains. This is suggestive of an unspecific export mechanism, such as cell lysis. Yet, also amongst the most abundantly detectable and invariant exoproteins of *P. gingivalis* there were proteins lacking signal peptides, which is suggestive of specific export, stable extracellular maintenance in the presence of gingipains, and a possible function in the bacterial life cycle. As to possible functions, it has been shown that proteins with important cytoplasmic functions, like elongation factors and proteins involved in central carbon metabolism, can serve important extracytoplasmic 'moonlighting' functions in bacterial adhesion to mammalian cells and tissues[133, 136, 137]. Altogether, it seems that ECP in Gram-negative bacteria may be more complex than initially thought, with several distinct pathways present[132]. Importantly, proteins subject to ECP can be predicted using SecretomeP 2.0[138].

Dedicated pipeline to approximate subcellular protein localization in *P. gingivalis*

To approximate subcellular protein localization in *P. gingivalis*, an in-house script implementing the decision tree presented in Figure 4 was developed. In addition to the aforementioned algorithms, TMHMM[139] was included to predict transmembrane helices, BOMP[140] to predict β-barrel OM proteins, and InterPro Scan[141] version 5.27 to detect particular domains in the InterPro consortium database[142] version 66. Based on *P. gingivalis* proteins of known location (Table S1), three lists of domains specific of PorSS cargo, IM proteins and OM proteins were devised (Table S2). Such domains, mainly structural, were chosen to be as specific as possible, in order to avoid biases. Using the software listed in Table 3, localization data were generated for all the seven revisited *P. gingivalis* strains. Further, following the flow scheme presented in Figure 4, a knowledge-based approach was implemented that is grounded on the currently available understanding of protein sorting systems active in *P. gingivalis* as detailed in the aforementioned sections. Importantly, the hierarchy of decisions in this pipeline was tailored to minimize

mistakes and biases, and to maximize compensation for possible software weaknesses.

Table 3. *Summary of predicted protein localizations*. Overview of the protein localization predictions for each strain enumerating all the proteins present in the different subcellular compartments.

ATCC 33277		W83		TDC60		MDS33		MDS140		512915		20655	
CYT	1286	CYT	1193	CYT	1451	CYT	1375	CYT	1362	CYT	1426	CYT	1366
ECP	95	ECP	80	ECP	109	ECP	94	ECP	97	ECP	103	ECP	115
IM	3	IM	3	IM	3	IM	3	IM	3	IM	3	IM	3
IM LP	18	IM LP	19	IM LP	18	IM LP	20	IM LP	22	IM LP	20	IM LP	19
IM TM	333	IM TM	316	IM TM	348	IM TM	332	IM TM	318	IM TM	342	IM TM	363
OM	57	OM	58	OM	60	OM	61	OM	56	OM	62	OM	64
OM LP	67	OM LP	46	OM LP	54	OM LP	54	OM LP	55	OM LP	61	OM LP	56
PERI	135	PERI	118	PERI	118	PERI	138	PERI	125	PERI	131	PERI	136
Secreted PorSS	22	Secreted PorSS	22	Secreted PorSS	24	Secreted PorSS	20	Secreted PorSS	21	Secreted PorSS	19	Secreted PorSS	20
UNK	6	UNK	8	UNK	9	UNK	9	UNK	6	UNK	5	UNK	5
Total	**2022**	**Total**	**1863**	**Total**	**2194**	**Total**	**2106**	**Total**	**2065**	**Total**	**2172**	**Total**	**2147**
TOT EXTRA (ECP + Secreted PorSS)	117	**TOT EXTRA** (ECP + Secreted PorSS)	102	**TOT EXTRA** (ECP + Secreted PorSS)	133	**TOT EXTRA** (ECP + Secreted PorSS)	114	**TOT EXTRA** (ECP + Secreted PorSS)	118	**TOT EXTRA** (ECP + Secreted PorSS)	122	**TOT EXTRA** (ECP + Secreted PorSS)	135
TOT IM (IM + IM LP + IM TM)	354	**TOT IM** (IM + IM LP + IM TM)	338	**TOT IM** (IM + IM LP + IM TM)	369	**TOT IM** (IM + IM LP + IM TM)	355	**TOT IM** (IM + IM LP + IM TM)	343	**TOT IM** (IM + IM LP + IM TM)	365	**TOT IM** (IM + IM LP + IM TM)	385
TOT OM (OM + OM LP)	124	**TOT OM** (OM + OM LP)	104	**TOT OM** (OM + OM LP)	114	**TOT OM** (OM + OM LP)	115	**TOT OM** (OM + OM LP)	111	**TOT OM** (OM + OM LP)	123	**TOT OM** (OM + OM LP)	120

Proteins displaying at least one of the selected PorSS cargo-specific domains (Table S2) were immediately designated as secreted *via* PorSS (Fig. 4), as these signatures are highly reliable in predicting secretion *via* this pathway. On this basis, the inspected *P. gingivalis* strains potentially secreted between 19 and 24 proteins specifically via the PorSS (Table 3). Despite its high specificity, this approach does not guarantee the identification of all PorSS cargo proteins, because some proteins exported *via* PorSS may lack the selected PorSS cargo domains. Further, the present listing of predicted PorSS

cargo proteins (Table S3) may represent an underestimation since potential misidentifications were not manually curated in order to avoid bias. This explains why the present number of potential PorSS cargo proteins is lower than the previously proposed 30 to 35 cargo proteins[76, 143], which may include some proteins whose secretion is indirectly related to the PorSS. In the second step of the prediction pipeline, both the LipoP and Lipo algorithms were used to identify lipoproteins amongst those proteins that were not assigned as PorSS cargo (Fig. 4). Since there was no possibility for a majority vote, the LipoP predictions were given priority in case of disagreement. The same approach was used to assess the signal peptidase II cleavage sites, which was necessary to pinpoint amino acid residues at positions +2 and +3 of the mature lipoproteins. In case of absence of a glycine residue at both positions, the protein was predicted to be an OM lipoprotein (OM LP). If a glycine residue was present at the +2 or +3 position, an additional control was performed by assessing the presence of a known OM domain. In the presence of an OM domain (Table S2) the G+2/+3 rule was ignored and the protein was still predicted as an OM lipoprotein (OM LP). Conversely, the lack of OM domains resulted in a protein's designation as an IM lipoprotein (IM LP). The numbers of predicted IM lipoproteins ranged from 18 to 22, and the numbers of OM lipoproteins ranged from 46 to 67 (Table 3). The large variation in the OM lipoproteins of different strains may relate to previously observed genomic rearrangements[144].

For non-lipoproteins, an agreed number of transmembrane helices as predicted by TMHMM and Phobius equal to, or higher than two was used to predict IM transmembrane proteins (Fig. 4). When instead both programs agreed on the presence of at least one transmembrane helix, a signal peptide check was performed, in order to reduce the number of false positively predicted helices caused by the signal peptide's presence. When the signal peptide prediction consensus was one or lower (i.e. one, or less, out of the three programs used predicted a signal peptide) the predicted helix was considered a 'true positive', and the respective protein was therefore predicted to have a transmembrane localization in the IM. In case of a signal peptide consensus equal or higher than two (i.e. at least two out of the three programs used predicted a signal peptide) and an agreed number of transmembrane helices as predicted by TMHMM and Phobius equal or lower than one, the protein was considered to have a signal peptide. In this case, it was further analyzed to determine a possible IM, OM, or periplasmic localization. Instead, when the signal peptide

consensus was one or zero, and the agreed number of transmembrane helices as predicted by TMHMM and Phobius was one or zero, the protein was considered to lack a signal peptide. Thus, despite being allegedly unable to cross the IM *via* Sec, such a protein was further analyzed for a possible cytoplasmic localization, or a potential IM, OM or extracellular localization *via* ECP. In all other cases, i.e. when the outcome of the predictions of transmembrane helices and the presence of a signal peptide were in conflict, the predicted localization of the respective protein was designated as unknown (Table 3). Merely five to nine proteins with unknown localization were encountered for the presently evaluated strains, suggesting that the approach adopted was extremely discriminative and robust.

Proteins with a signal peptide are able to cross the IM, ending up in the periplasm. The additional presence of either a β-barrel, indicated by a BOMP score higher than three, or of at least one OM domain (Table S2), can be considered as an indicator for subsequent association with or insertion into the OM. The latter proteins were thus predicted to localize in the OM compartment. In case a TM domain, β-barrel or OM domain were absent, while a signal peptide was present, the respective protein was designated as having a periplasmic localization. Since the presence of a Phobius-predicted TM domain is indicative of protein retention in the IM, such proteins were designated as IM resident-proteins.

In the absence of a signal peptide, a protein can be retained in the cytosol, be secreted through non-canonical or unknown pathways thus ending in the OM or extracellular milieu, or still be inserted into or associated with the IM. For such reasons, all proteins lacking a predicted signal peptide were checked for the presence of OM domains (Table S2). If one or more of such domains were present, the respective protein was designated to have an OM localization. Analogously, the presence of an IM-related domain (Table S2) was used as an indicator for IM localization.

A SecretomeP analysis was performed as the last verification step in the prediction pipeline (Fig. 4), because of its knowledge-based nature. At this juncture, the remaining proteins exhibit no relevant feature as discussed above and, accordingly, their sorting destination could only be the cytoplasm or the extracellular milieu due to non-canonical or unknown ECP pathways. Therefore, in the presence of a SecretomeP score equal or higher than 0.75, proteins were predicted to undergo ECP. A lower score, instead, pointed at a cytosolic

localization, since none of the applied predictors suggested the possibility of the respective protein leaving the cytosol. The overall outcome of the predicted protein localization in *P. gingivalis* is listed in Table S3, while Table 3 presents an overview of these predictions.

Core and variant exoproteome analyses

Interestingly, analysis of the *P. gingivalis* exoproteome highlighted strain-specific variations[49], which were also encountered in the present inspection of subcellular protein localization. This was the incentive for a bioinformatics-based appraisal of the core and variant (exo)proteome of *P. gingivalis*. Thus, to identify orthologs in the proteome of different strains, reciprocal best hits (RBHs) were calculated. In brief, Galaxy[145] was used to perform reciprocal protein BLAST searches (NCBI BLAST+ v. 2.3.0[146]). Default parameters (minimum percentage identity: 70%; minimum High Scoring Pair (HSP) coverage: 50%) were used and all redundancies were removed prior the BLAST search. RBHs were then calculated by blasting the deduced amino acid sequences of all investigated strains against those of *P. gingivalis* ATCC 33277. Despite *P. gingivalis* W83 being the most used reference strain in the field, the ATCC 33277 strain was adopted as a reference for the present analyses after the realization that many proteins were actually encoded by the W83 genome sequence while the respective genes were never annotated (data not shown). Tblastn was used to identify some of these proteins, being part of the main secretion complexes, and they are reported in Table 1. The core proteome was thus defined by the set of proteins having an ortholog in all six strains analyzed against the ATCC 33277 reference strain (Table S4). The remaining protein complement identified for each strain is regarded as the respective variable proteome (Table S5). Of note, considering possible misannotations of the used genome sequences, the presently proposed distinction between the *P. gingivalis* core and variable proteomes should be regarded as an approximation rather than an absolute distinction.

To predict the core exoproteome, the proteins in the core proteome were divided according to their possible subcellular localizations, as per our prediction pipeline, and two categories were pulled together: 1) proteins of the OM compartment (OM LP and OM proteins) and 2) PorSS cargo proteins (Table S6). The GO terms associated with the domains detected by InterPro for these exoproteins were taken into

account for each strain. The obtained GO terms were then used in a REVIGO[147] analysis, to unravel the network of biological pathways created by the core exoproteome. It should be noted that the potential ECP complement as designated by our pipeline was excluded from the exoproteome classification due to its high variability between strains (Table S3). The remaining predicted exoproteins are, instead, almost entirely predicted to make up the core exoproteome. The scarce suitability of SecretomeP for our *P. gingivalis* dataset is probably due to the fact that ECP predictors cannot be tailored on specific transport systems or bacterial species due to their intrinsic nature. GO term analysis of the core exoproteome predicted for each inspected strain yielded close to identical results (Fig. 5) with some marginal differences observed for strain MDS33 (data not shown). The latter may relate to minor discrepancies in the genome annotation of strain MDS33, or to some small potential differences in the core orthologs of this strain. The identified core exoproteins operated in eight different major biological pathways, namely putrescine biosynthesis, intracellular protein transport, membrane assembly, protein folding, metabolism, carbohydrate metabolism, proteolysis, and oxidation-reduction processes (Fig. 5).

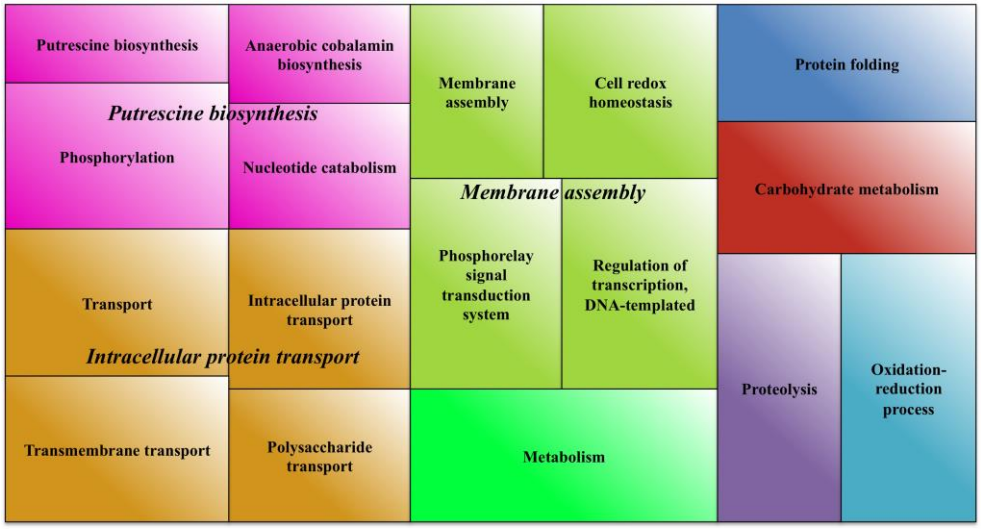

Figure 5. *Biological pathways represented in the P. gingivalis core exoproteome.* REVIGO treemap depicting the outcome of a GO term analysis of cellular pathways involving the proteins predicted to define the core exoproteome of the *P. gingivalis* strains under examination.

As these results slightly differed from previous observations on the core exoproteome of a different and smaller set of samples[49], mainly for the lack of a pathogenesis GO term cluster, we also analyzed the variable exoproteome (Table S7). The simple absence of one virulence factor from one strain, in fact, would eliminate the protein from the core exoproteome and relegate it to the variable exoproteome. As expected, the GO term analyses of the variable exoproteomes revealed a sizable amount of extracellular proteins involved in pathogenesis in all *P. gingivalis* strains, except MDS140. The latter strain happens to be isolated from a healthy carrier. It is therefore tempting to speculate that the MDS140 strain could lack a number of virulence factors. Additionally, only the 'transport' and 'proteolysis' labels were assigned to predicted exoproteins of the MDS140 isolate (Fig. S1A), in contrast to the various other functional labels assigned to exoproteins from the other investigated strains (Fig. S1B-D).
Lastly, all the protein sorting information gathered by reviewing the available literature and predicting subcellular protein localization in *P. gingivalis* has been combined in Figure 6, which presents the total numbers of proteins predicted per subcellular compartment.

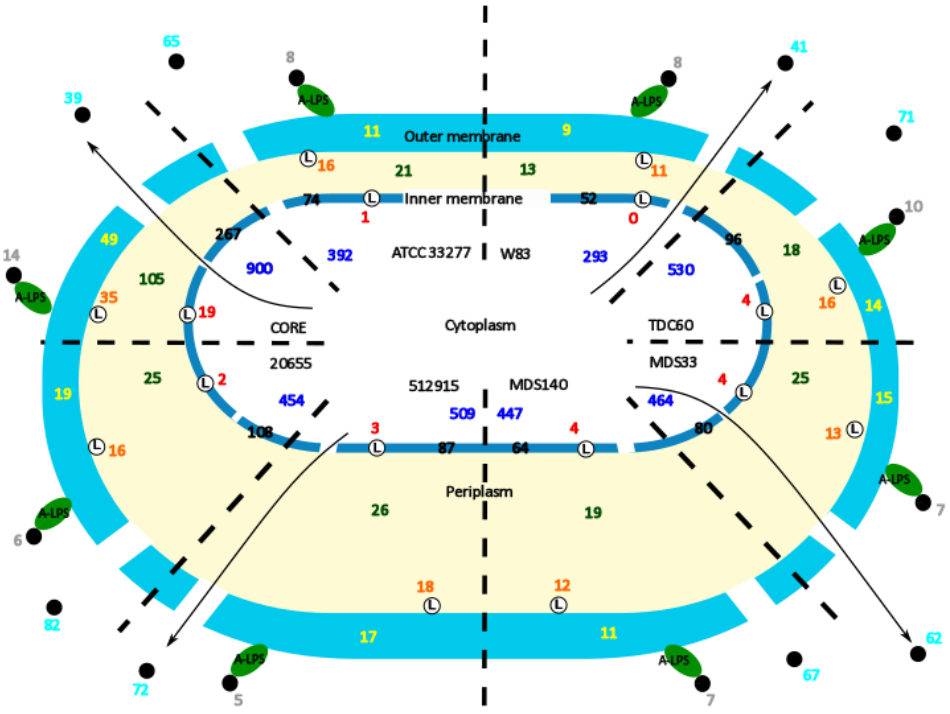

Figure 6. *Overview of the subcellular localization of core and variant P. gingivalis proteins*. The numbers of proteins residing at a particular subcellular

localization or extracellularly are indicated for the core proteome and the variable proteome of each *P. gingivalis* strain under examination. In blue the amount of cytoplasmic proteins, red for IM lipoproteins, black for IM proteins, green for periplasmic proteins, orange for OM lipoproteins, yellow for OM proteins, grey for PorSS secreted proteins and cyan for ECP-secreted proteins.

Of note, this overview image distinguishes the core and variant proteomes of each compartment.

Conclusion

This review is focused on protein localization in the oral pathogen *P. gingivalis* integrating the results of published biochemical studies[47, 50-52] and a tailored *in silico* evaluation of published genome sequences grounded on established bioinformatic approaches[117, 118, 125]. Considering the broad spectrum of interests that *P. gingivalis* elicits, especially in the fields of periodontology, rheumatology and microbiology, this review will serve as an important lead for many upcoming studies concerning this bacterium. In fact, a compendium of the different subcellular and extracellular destination(s) that each individual protein may reach constitutes a treasure trove of invaluable information for any kind of research involving the biology and virulence of this bacterium. A sterling example of this is the importance of the exoproteome for virulence, adhesion, diagnostic, clinical, and biofilm development studies. This bacterial road map could therefore be used to devise diagnostic or therapeutic antibodies targeting specific surface proteins, to create vaccines, and to discover druggable targets. Additionally, as several proteins of *P. gingivalis* are subject of ongoing studies, the data regarding the proteins belonging to the same subcellular compartments is a significant advantage when looking for targets, inhibitors or cofactors. A simple example of this is utilizing exoproteomic data to narrow down the list of possible targets of the citrullinating PPAD enzyme.
Lastly, *P. gingivalis* is an extremely successful oral pathogen that takes advantage of several virulence factors, such as capsule[148, 149], cysteine proteinases, fimbriae, lipopolysaccharide[150] and a unique peptidylarginine deiminase to manifest itself as a “keystone” species within subgingival biofilms[151, 152]. All the mechanisms contributing to the success of this bacterium in the periodontal pockets, where it mainly resides, are contained in its proteome. Hence understanding the mechanisms and magnitude of protein localization events in *P.*

gingivalis will be essential to avoid or avert the venomous bite of this oral pathogen.

Competing interests

The authors declare that they have no financial and non-financial competing interests in relation to the documented research.

References

1. van Winkelhoff, A. J., Loos, B. G., van der Reijden, W. A. & van der Velden, U. *Porphyromonas gingivalis, Bacteroides forsythus* and other putative periodontal pathogens in subjects with and without periodontal destruction. *J. Clin. Periodontol.* **29**, 1023-1028 (2002).
2. Bostanci, N. & Belibasakis, G. N. *Porphyromonas gingivalis*: an invasive and evasive opportunistic oral pathogen. *FEMS Microbiol. Lett.* **333**, 1-9 (2012).
3. Yang, H. W., Huang, Y. F. & Chou, M. Y. Occurrence of *Porphyromonas gingivalis* and *Tannerella forsythensis* in periodontally diseased and healthy subjects. *J. Periodontol.* **75**, 1077-1083 (2004).
4. Rylev, M. & Kilian, M. Prevalence and distribution of principal periodontal pathogens worldwide. *J. Clin. Periodontol.* **35**, 346-361 (2008).
5. Darveau, R. P. Periodontitis: a polymicrobial disruption of host homeostasis. *Nat. Rev. Microbiol.* **8**, 481-490 (2010).
6. Potempa, J., Mydel, P. & Koziel, J. The case for periodontitis in the pathogenesis of rheumatoid arthritis. *Nat. Rev. Rheumatol.* (2017).
7. Zhu, M. & Nikolajczyk, B. S. Immune cells link obesity-associated type 2 diabetes and periodontitis. *J. Dent. Res.* **93**, 346-352 (2014).
8. Chee, B., Park, B. & Bartold, P. M. Periodontitis and type II diabetes: a two-way relationship. *Int. J. Evid Based. Healthc.* **11**, 317-329 (2013).
9. Preshaw, P. M. & Bissett, S. M. Periodontitis: oral complication of diabetes. *Endocrinol. Metab. Clin. North Am.* **42**, 849-867 (2013).
10. Kumar, M., Mishra, L., Mohanty, R. & Nayak, R. "Diabetes and gum disease: the diabolic duo". *Diabetes Metab. Syndr.* **8**, 255-258 (2014).

11. Bascones-Martinez, A., Gonzalez-Febles, J. & Sanz-Esporrin, J. Diabetes and periodontal disease. Review of the literature. *Am. J. Dent.* **27**, 63-67 (2014).
12. Gurav, A. N. The association of periodontitis and metabolic syndrome. *Dent. Res. J. (Isfahan)* **11**, 1-10 (2014).
13. Kebschull, M., Demmer, R. T. & Papapanou, P. N. "Gum bug, leave my heart alone!"--epidemiologic and mechanistic evidence linking periodontal infections and atherosclerosis. *J. Dent. Res.* **89**, 879-902 (2010).
14. Kelly, J. T., Avila-Ortiz, G., Allareddy, V., Johnson, G. K. & Elangovan, S. The association between periodontitis and coronary heart disease: a quality assessment of systematic reviews. *J. Am. Dent. Assoc.* **144**, 371-379 (2013).
15. Li, C. *et al.* Periodontal therapy for the management of cardiovascular disease in patients with chronic periodontitis. *Cochrane Database Syst. Rev.* **8**, CD009197 (2014).
16. Noble, J. M., Scarmeas, N. & Papapanou, P. N. Poor oral health as a chronic, potentially modifiable dementia risk factor: review of the literature. *Curr. Neurol. Neurosci. Rep.* **13**, 384-013-0384-x (2013).
17. Shaik, M. M. *et al.* How do periodontal infections affect the onset and progression of Alzheimer's disease? *CNS Neurol. Disord. Drug Targets* **13**, 460-466 (2014).
18. Kamer, A. R. *et al.* Inflammation and Alzheimer's disease: possible role of periodontal diseases. *Alzheimers Dement.* **4**, 242-250 (2008).
19. Kamer, A. R. *et al.* Alzheimer's disease and peripheral infections: the possible contribution from periodontal infections, model and hypothesis. *J. Alzheimers Dis.* **13**, 437-449 (2008).
20. de Pablo, P., Dietrich, T. & McAlindon, T. E. Association of periodontal disease and tooth loss with rheumatoid arthritis in the US population. *J. Rheumatol.* **35**, 70-76 (2008).
21. de Smit, M. *et al.* Periodontitis in established rheumatoid arthritis patients: a cross-sectional clinical, microbiological and serological study. *Arthritis Res. Ther.* **14**, R222 (2012).
22. de Smit, M. J. *et al.* Effect of periodontal treatment on rheumatoid arthritis and vice versa. *Ned. Tijdschr. Tandheelkd.* **119**, 191-197 (2012).
23. de Smit, M. J., Brouwer, E., Vissink, A. & van Winkelhoff, A. J. Rheumatoid arthritis and periodontitis; a possible link via citrullination. *Anaerobe* **17**, 196-200 (2011).

24. Detert, J., Pischon, N., Burmester, G. R. & Buttgereit, F. The association between rheumatoid arthritis and periodontal disease. *Arthritis Res. Ther.* **12**, 218 (2010).
25. Dissick, A. *et al.* Association of periodontitis with rheumatoid arthritis: a pilot study. *J. Periodontol.* **81**, 223-230 (2010).
26. El-Shinnawi, U. & Soory, M. Associations between periodontitis and systemic inflammatory diseases: response to treatment. *Recent. Pat. Endocr Metab. Immune Drug Discov.* **7**, 169-188 (2013).
27. Farquharson, D., Butcher, J. P. & Culshaw, S. Periodontitis, *Porphyromonas*, and the pathogenesis of rheumatoid arthritis. *Mucosal Immunol.* **5**, 112-120 (2012).
28. Hitchon, C. A. & El-Gabalawy, H. S. Infection and rheumatoid arthritis: still an open question. *Curr. Opin. Rheumatol.* **23**, 352-357 (2011).
29. Lundberg, K., Wegner, N., Yucel-Lindberg, T. & Venables, P. J. Periodontitis in RA-the citrullinated enolase connection. *Nat. Rev. Rheumatol.* **6**, 727-730 (2010).
30. Maresz, K. J. *et al. Porphyromonas gingivalis* facilitates the development and progression of destructive arthritis through its unique bacterial peptidylarginine deiminase (PAD). *PLoS Pathog.* **9**, e1003627 (2013).
31. Mercado, F., Marshall, R. I., Klestov, A. C. & Bartold, P. M. Is there a relationship between rheumatoid arthritis and periodontal disease? *J. Clin. Periodontol.* **27**, 267-272 (2000).
32. Mercado, F. B., Marshall, R. I., Klestov, A. C. & Bartold, P. M. Relationship between rheumatoid arthritis and periodontitis. *J. Periodontol.* **72**, 779-787 (2001).
33. Pischon, N. *et al.* Association among rheumatoid arthritis, oral hygiene, and periodontitis. *J. Periodontol.* **79**, 979-986 (2008).
34. Quirke, A. M. *et al.* Heightened immune response to autocitrullinated *Porphyromonas gingivalis* peptidylarginine deiminase: a potential mechanism for breaching immunologic tolerance in rheumatoid arthritis. *Ann. Rheum. Dis.* **73**, 263-269 (2014).
35. Routsias, J. G., Goules, J. D., Goules, A., Charalampakis, G. & Pikazis, D. Autopathogenic correlation of periodontitis and rheumatoid arthritis. *Rheumatology (Oxford)* **50**, 1189-1193 (2011).
36. Tolo, K. & Jorkjend, L. Serum antibodies and loss of periodontal bone in patients with rheumatoid arthritis. *J. Clin. Periodontol.* **17**, 288-291 (1990).

37. Wegner, N. *et al.* Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. *Immunol. Rev.* **233**, 34-54 (2010).
38. Wegner, N. *et al.* Peptidylarginine deiminase from *Porphyromonas gingivalis* citrullinates human fibrinogen and alpha-enolase: implications for autoimmunity in rheumatoid arthritis. *Arthritis Rheum.* **62**, 2662-2672 (2010).
39. Kasser, U. R. *et al.* Risk for periodontal disease in patients with longstanding rheumatoid arthritis. *Arthritis Rheum.* **40**, 2248-2251 (1997).
40. Nesse, W. *et al.* Increased prevalence of cardiovascular and autoimmune diseases in periodontitis patients: a cross-sectional study. *J. Periodontol.* **81**, 1622-1628 (2010).
41. Mangat, P., Wegner, N., Venables, P. J. & Potempa, J. Bacterial and human peptidylarginine deiminases: targets for inhibiting the autoimmune response in rheumatoid arthritis? *Arthritis Res. Ther.* **12**, 209 (2010).
42. Gabarrini, G. *et al.* The peptidylarginine deiminase gene is a conserved feature of *Porphyromonas gingivalis*. *Sci. Rep.* **5**, 13936 (2015).
43. Gabarrini, G. *et al.* Conserved Citrullinating Exoenzymes in *Porphyromonas* Species. *J. Dent. Res.*, 22034517747575 (2017).
44. Goulas, T. *et al.* Structure and mechanism of a bacterial host-protein citrullinating virulence factor, *Porphyromonas gingivalis* peptidylarginine deiminase. *Sci. Rep.* **5**, 11969 (2015).
45. Avouac, J., Gossec, L. & Dougados, M. Diagnostic and predictive value of anti-cyclic citrullinated protein antibodies in rheumatoid arthritis: a systematic literature review. *Ann. Rheum. Dis.* **65**, 845-851 (2006).
46. Nishimura, K. *et al.* Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. *Ann. Intern. Med.* **146**, 797-808 (2007).
47. Sato, K. *et al.* Identification of *Porphyromonas gingivalis* proteins secreted by the Por secretion system. *FEMS Microbiol. Lett.* **338**, 68-76 (2013).
48. Gabarrini, G. *et al.* There's no place like OM: Vesicular sorting and secretion of the peptidylarginine deiminase of *Porphyromonas gingivalis*. *Virulence* **9**, 456-464 (2018).

49. Stobernack, T. *et al.* Extracellular Proteome and Citrullinome of the Oral Pathogen *Porphyromonas gingivalis*. *J. Proteome Res.* **15**, 4532-4543 (2016).
50. Veith, P. D. *et al.* Major outer membrane proteins and proteolytic processing of RgpA and Kgp of *Porphyromonas gingivalis* W50. *Biochem. J.* **363**, 105-115 (2002).
51. Yoshimura, F., Murakami, Y., Nishikawa, K., Hasegawa, Y. & Kawaminami, S. Surface components of *Porphyromonas gingivalis*. *J. Periodontal. Res.* **44**, 1-12 (2009).
52. McBride, M. J. & Zhu, Y. Gliding motility and Por secretion system genes are widespread among members of the phylum bacteroidetes. *J. Bacteriol.* **195**, 270-278 (2013).
53. Scheres, N., Lamont, R. J., Crielaard, W. & Krom, B. P. LuxS signaling in *Porphyromonas gingivalis*-host interactions. *Anaerobe* **35**, 3-9 (2015).
54. Glew, M. D. *et al.* Blue native-PAGE analysis of membrane protein complexes in *Porphyromonas gingivalis*. *J. Proteomics* **110**, 72-92 (2014).
55. Rangarajan, M. *et al.* LptO (PG0027) Is Required for Lipid A 1-Phosphatase Activity in *Porphyromonas gingivalis* W50. *J. Bacteriol.* **199**, 10.1128/JB.00751-16. Print 2017 Jun 1 (2017).
56. Nonaka, M. *et al.* Analysis of a Lys-specific serine endopeptidase secreted via the type IX secretion system in *Porphyromonas gingivalis*. *FEMS Microbiol. Lett.* **354**, 60-68 (2014).
57. Saiki, K. & Konishi, K. *Porphyromonas gingivalis* C-terminal signal peptidase PG0026 and HagA interact with outer membrane protein PG27/LptO. *Mol. Oral Microbiol.* **29**, 32-44 (2014).
58. Hasegawa, Y. *et al.* Localization and function of the accessory protein Mfa3 in *Porphyromonas gingivalis* Mfa1 fimbriae. *Mol. Oral Microbiol.* **28**, 467-480 (2013).
59. Mantri, C. K. *et al.* Fimbriae-mediated outer membrane vesicle production and invasion of *Porphyromonas gingivalis*. *Microbiologyopen* **4**, 53-65 (2015).
60. Nagano, K., Hasegawa, Y., Yoshida, Y. & Yoshimura, F. A Major Fimbrilin Variant of Mfa1 Fimbriae in *Porphyromonas gingivalis*. *J. Dent. Res.* **94**, 1143-1148 (2015).
61. Staniec, D. *et al.* Calcium Regulates the Activity and Structural Stability of Tpr, a Bacterial Calpain-like Peptidase. *J. Biol. Chem.* **290**, 27248-27260 (2015).

62. Gorasia, D. G. *et al.* *Porphyromonas gingivalis* Type IX Secretion Substrates Are Cleaved and Modified by a Sortase-Like Mechanism. *PLoS Pathog.* **11**, e1005152 (2015).
63. Ota, K. *et al.* SigCH, an extracytoplasmic function sigma factor of *Porphyromonas gingivalis* regulates the expression of cdhR and hmuYR. *Anaerobe* **43**, 82-90 (2017).
64. Lasica, A. M. *et al.* Structural and functional probing of PorZ, an essential bacterial surface component of the type-IX secretion system of human oral-microbiomic *Porphyromonas gingivalis*. *Sci. Rep.* **6**, 37708 (2016).
65. Heath, J. E. *et al.* PG1058 Is a Novel Multidomain Protein Component of the Bacterial Type IX Secretion System. *PLoS One* **11**, e0164313 (2016).
66. Goulas, T. *et al.* Structure of RagB, a major immunodominant outer-membrane surface receptor antigen of *Porphyromonas gingivalis*. *Mol. Oral Microbiol.* **31**, 472-485 (2016).
67. Naylor, K. L. *et al.* Role of OmpA2 surface regions of *Porphyromonas gingivalis* in host-pathogen interactions with oral epithelial cells. *Microbiologyopen* **6**, 10.1002/mbo3.401. Epub 2016 Sep 6 (2017).
68. Gorasia, D. G. *et al.* Structural Insights into the PorK and PorN Components of the *Porphyromonas gingivalis* Type IX Secretion System. *PLoS Pathog.* **12**, e1005820 (2016).
69. Hasegawa, Y. *et al.* Role of Mfa5 in Expression of Mfa1 Fimbriae in *Porphyromonas gingivalis*. *J. Dent. Res.* **95**, 1291-1297 (2016).
70. Smalley, J. W. & Olczak, T. Heme acquisition mechanisms of *Porphyromonas gingivalis* - strategies used in a polymicrobial community in a heme-limited host environment. *Mol. Oral Microbiol.* **32**, 1-23 (2017).
71. Taguchi, Y. *et al.* Involvement of an Skp-Like Protein, PGN_0300, in the Type IX Secretion System of *Porphyromonas gingivalis*. *Infect. Immun.* **84**, 230-240 (2015).
72. Slakeski, N. *et al.* A *Porphyromonas gingivalis* genetic locus encoding a heme transport system. *Oral Microbiol. Immunol.* **15**, 388-392 (2000).
73. Karunakaran, T., Madden, T. & Kuramitsu, H. Isolation and characterization of a hemin-regulated gene, hemR, from *Porphyromonas gingivalis*. *J. Bacteriol.* **179**, 1898-1908 (1997).
74. Fujise, K., Kikuchi, Y., Kokubu, E., Okamoto-Shibayama, K. & Ishihara, K. Effect of extracytoplasmic function sigma factors on

autoaggregation, hemagglutination, and cell surface properties of *Porphyromonas gingivalis*. *PLoS One* **12**, e0185027 (2017).
75. Veith, P. D., Glew, M. D., Gorasia, D. G. & Reynolds, E. C. Type IX secretion: the generation of bacterial cell surface coatings involved in virulence, gliding motility and the degradation of complex biopolymers. *Mol. Microbiol.* **106**, 35-53 (2017).
76. Lasica, A. M., Ksiazek, M., Madej, M. & Potempa, J. The Type IX Secretion System (T9SS): Highlights and Recent Insights into Its Structure and Function. *Front. Cell. Infect. Microbiol.* **7**, 215 (2017).
77. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158-D169 (2017).
78. Hoover, C. I., Abarbarchuk, E., Ng, C. Y. & Felton, J. R. Transposition of Tn4351 in *Porphyromonas gingivalis*. *Plasmid* **27**, 246-250 (1992).
79. Sandmeier, H., Bar, K. & Meyer, J. Search for bacteriophages of black-pigmented gram-negative anaerobes from dental plaque. *FEMS Immunol. Med. Microbiol.* **6**, 193-194 (1993).
80. Veith, P. D. *et al. Porphyromonas gingivalis* outer membrane vesicles exclusively contain outer membrane and periplasmic proteins and carry a cargo enriched with virulence factors. *J. Proteome Res.* **13**, 2420-2432 (2014).
81. Gui, M. J., Dashper, S. G., Slakeski, N., Chen, Y. Y. & Reynolds, E. C. Spheres of influence: *Porphyromonas gingivalis* outer membrane vesicles. *Mol. Oral Microbiol.* **31**, 365-378 (2016).
82. Ellis, T. N. & Kuehn, M. J. Virulence and immunomodulatory roles of bacterial outer membrane vesicles. *Microbiol. Mol. Biol. Rev.* **74**, 81-94 (2010).
83. O'Brien-Simpson, N. M., Pathirana, R. D., Walker, G. D. & Reynolds, E. C. *Porphyromonas gingivalis* RgpA-Kgp proteinase-adhesin complexes penetrate gingival tissue and induce proinflammatory cytokines or apoptosis in a concentration-dependent manner. *Infect. Immun.* **77**, 1246-1261 (2009).
84. Furuta, N., Takeuchi, H. & Amano, A. Entry of *Porphyromonas gingivalis* outer membrane vesicles into epithelial cells causes cellular functional impairment. *Infect. Immun.* **77**, 4761-4770 (2009).
85. Furuta, N. *et al. Porphyromonas gingivalis* outer membrane vesicles enter human epithelial cells via an endocytic pathway and are sorted to lysosomal compartments. *Infect. Immun.* **77**, 4187-4196 (2009).

86. Zhu, Y. *et al*. *Porphyromonas gingivalis* and *Treponema denticola* synergistic polymicrobial biofilm development. *PLoS One* **8**, e71727 (2013).
87. O-Brien-Simpson, N. M., Veith, P. D., Dashper, S. G. & Reynolds, E. C. *Porphyromonas gingivalis* gingipains: the molecular teeth of a microbial vampire. *Curr. Protein Pept. Sci.* **4**, 409-426 (2003).
88. Potempa, J., Banbula, A. & Travis, J. Role of bacterial proteinases in matrix destruction and modulation of host responses. *Periodontol. 2000* **24**, 153-192 (2000).
89. Driessen, A. J. & Nouwen, N. Protein translocation across the bacterial cytoplasmic membrane. *Annu. Rev. Biochem.* **77**, 643-667 (2008).
90. Sargent, F. The twin-arginine transport system: moving folded proteins across membranes. *Biochem. Soc. Trans.* **35**, 835-847 (2007).
91. Robinson, C. *et al*. Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria. *Biochim. Biophys. Acta* **1808**, 876-884 (2011).
92. Goosens, V. J. & van Dijl, J. M. Twin-Arginine Protein Translocation. *Curr. Top. Microbiol. Immunol.* **404**, 69-94 (2017).
93. Knowles, T. J., Scott-Tucker, A., Overduin, M. & Henderson, I. R. Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nat. Rev. Microbiol.* **7**, 206-214 (2009).
94. Konovalova, A., Kahne, D. E. & Silhavy, T. J. Outer Membrane Biogenesis. *Annu. Rev. Microbiol.* **71**, 539-556 (2017).
95. Narita, S. & Tokuda, H. An ABC transporter mediating the membrane detachment of bacterial lipoproteins depending on their sorting signals. *FEBS Lett.* **580**, 1164-1170 (2006).
96. Okuda, S. & Tokuda, H. Lipoprotein sorting in bacteria. *Annu. Rev. Microbiol.* **65**, 239-259 (2011).
97. Rossiter, A. E. *et al*. The essential beta-barrel assembly machinery complex components BamD and BamA are required for autotransporter biogenesis. *J. Bacteriol.* **193**, 4250-4253 (2011).
98. Desvaux, M., Hebraud, M., Talon, R. & Henderson, I. R. Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends Microbiol.* **17**, 139-145 (2009).
99. Low, H. H. *et al*. Structure of a type IV secretion system. *Nature* **508**, 550-553 (2014).

100. Leone, P. *et al.* Type IX secretion system PorM and gliding machinery GldM form arches spanning the periplasmic space. *Nat. Commun.* **9**, 429-017-02784-7 (2018).
101. Shoji, M. *et al.* Por secretion system-dependent secretion and glycosylation of *Porphyromonas gingivalis* hemin-binding protein 35. *PLoS One* **6**, e21372 (2011).
102. Konig, M. F., Paracha, A. S., Moni, M., Bingham, C. O.,3rd & Andrade, F. Defining the role of *Porphyromonas gingivalis* peptidylarginine deiminase (PPAD) in rheumatoid arthritis through the study of PPAD biology. *Ann. Rheum. Dis.* **74**, 2054-2061 (2015).
103. Eichinger, A. *et al.* Crystal structure of gingipain R: an Arg-specific bacterial cysteine proteinase with a caspase-like fold. *EMBO J.* **18**, 5453-5462 (1999).
104. Chen, Y. Y. *et al.* CPG70 is a novel basic metallocarboxypeptidase with C-terminal polycystic kidney disease domains from *Porphyromonas gingivalis*. *J. Biol. Chem.* **277**, 23433-23440 (2002).
105. Mikolajczyk, J. *et al.* Sequential autolytic processing activates the zymogen of Arg-gingipain. *J. Biol. Chem.* **278**, 10458-10464 (2003).
106. Ngo, L. H. *et al.* Mass spectrometric analyses of peptides and proteins in human gingival crevicular fluid. *J. Proteome Res.* **9**, 1683-1693 (2010).
107. Paramonov, N. *et al.* Structural analysis of the polysaccharide from the lipopolysaccharide of *Porphyromonas gingivalis* strain W50. *Eur. J. Biochem.* **268**, 4698-4707 (2001).
108. Xie, H. Biogenesis and function of *Porphyromonas gingivalis* outer membrane vesicles. *Future Microbiol.* **10**, 1517-1527 (2015).
109. Gabarrini, G. *et al.* Dropping anchor: attachment of peptidylarginine deiminase via A-LPS to secreted outer membrane vesicles of *Porphyromonas gingivalis*. *Sci. Rep.* **8**, 8949-018-27223-5 (2018).
110. Rawlings, N. D. A large and accurate collection of peptidase cleavages in the MEROPS database. *Database (Oxford)* **2009**, bap015 (2009).
111. Dalbey, R. E., Wang, P. & van Dijl, J. M. Membrane proteases in the bacterial protein secretion and quality control pathway. *Microbiol. Mol. Biol. Rev.* **76**, 311-330 (2012).
112. van Roosmalen, M. L. *et al.* Type I signal peptidases of Gram-positive bacteria. *Biochim. Biophys. Acta* **1694**, 279-297 (2004).

113. Bochtler, M. *et al.* The Bacteroidetes Q-Rule: Pyroglutamate in Signal Peptidase I Substrates. *Front. Microbiol.* **9**, 230 (2018).
114. Nguyen, M. T. *et al.* Lipid moieties on lipoproteins of commensal and non-commensal staphylococci induce differential immune responses. *Nat. Commun.* **8**, 2246-017-02234-4 (2017).
115. Tjalsma, H., Bolhuis, A., Jongbloed, J. D., Bron, S. & van Dijl, J. M. Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol. Mol. Biol. Rev.* **64**, 515-547 (2000).
116. Antelmann, H. *et al.* A proteomic view on genome-based signal peptide predictions. *Genome Res.* **11**, 1484-1502 (2001).
117. Tjalsma, H. *et al.* Proteomics of protein secretion by *Bacillus subtilis*: separating the "secrets" of the secretome. *Microbiol. Mol. Biol. Rev.* **68**, 207-233 (2004).
118. Sibbald, M. J. *et al.* Mapping the pathways to staphylococcal pathogenesis by comparative secretomics. *Microbiol. Mol. Biol. Rev.* **70**, 755-788 (2006).
119. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785-786 (2011).
120. Hiller, K., Grote, A., Scheer, M., Munch, R. & Jahn, D. PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res.* **32**, W375-9 (2004).
121. Kall, L., Krogh, A. & Sonnhammer, E. L. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027-1036 (2004).
122. Juncker, A. S. *et al.* Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* **12**, 1652-1662 (2003).
123. Berven, F. S. *et al.* Analysing the outer membrane subproteome of *Methylococcus capsulatus* (Bath) using proteomics and novel biocomputing tools. *Arch. Microbiol.* **184**, 362-377 (2006).
124. Lewenza, S., Gardy, J. L., Brinkman, F. S. & Hancock, R. E. Genome-wide identification of *Pseudomonas aeruginosa* exported proteins using a consensus computational strategy combined with a laboratory-based PhoA fusion screen. *Genome Res.* **15**, 321-329 (2005).
125. Romine, M. F. Genome-wide protein localization prediction strategies for gram negative bacteria. *BMC Genomics* **12 Suppl 1**, S1-2164-12-S1-S1 (2011).

126. Tjalsma, H. & van Dijl, J. M. Proteomics-based consensus prediction of protein retention in a bacterial membrane. *Proteomics* **5**, 4472-4482 (2005).
127. Gennity, J. M. & Inouye, M. The protein sequence responsible for lipoprotein membrane localization in Escherichia coli exhibits remarkable specificity. *J. Biol. Chem.* **266**, 16458-16464 (1991).
128. Narita, S. & Tokuda, H. Amino acids at positions 3 and 4 determine the membrane specificity of *Pseudomonas aeruginosa* lipoproteins. *J. Biol. Chem.* **282**, 13372-13378 (2007).
129. Schulze, R. J. & Zuckert, W. R. *Borrelia burgdorferi* lipoproteins are secreted to the outer surface by default. *Mol. Microbiol.* **59**, 1473-1484 (2006).
130. Wiker, H. G., Wilson, M. A. & Schoolnik, G. K. Extracytoplasmic proteins of *Mycobacterium tuberculosis* - mature secreted proteins often start with aspartic acid and proline. *Microbiology* **146 (Pt 7)**, 1525-1533 (2000).
131. Tjalsma, H. *et al.* The role of lipoprotein processing by signal peptidase II in the Gram-positive eubacterium *Bacillus subtilis*. Signal peptidase II is required for the efficient secretion of alpha-amylase, a non-lipoprotein. *J. Biol. Chem.* **274**, 1698-1707 (1999).
132. Wang, G. *et al.* How are the non-classically secreted bacterial proteins released into the extracellular milieu? *Curr. Microbiol.* **67**, 688-695 (2013).
133. Ebner, P. *et al.* Non-classical Protein Excretion Is Boosted by PSMalpha-Induced Cell Leakage. *Cell. Rep.* **20**, 1278-1286 (2017).
134. Krishnappa, L. *et al.* Extracytoplasmic proteases determining the cleavage and release of secreted proteins, lipoproteins, and membrane proteins in *Bacillus subtilis*. *J. Proteome Res.* **12**, 4101-4110 (2013).
135. Berks, B. C. A common export pathway for proteins binding complex redox cofactors? *Mol. Microbiol.* **22**, 393-404 (1996).
136. Jeffery, C. J. Moonlighting proteins. *Trends Biochem. Sci.* **24**, 8-11 (1999).
137. Ebner, P. *et al.* Excreted Cytoplasmic Proteins Contribute to Pathogenicity in *Staphylococcus aureus*. *Infect. Immun.* **84**, 1672-1681 (2016).
138. Bendtsen, J. D., Kiemer, L., Fausboll, A. & Brunak, S. Non-classical protein secretion in bacteria. *BMC Microbiol.* **5**, 58 (2005).
139. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov

model: application to complete genomes. *J. Mol. Biol.* **305**, 567-580 (2001).
140. Berven, F. S., Flikka, K., Jensen, H. B. & Eidhammer, I. BOMP: a program to predict integral beta-barrel outer membrane proteins encoded within genomes of Gram-negative bacteria. *Nucleic Acids Res.* **32**, W394-9 (2004).
141. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).
142. Finn, R. D. *et al.* InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190-D199 (2017).
143. Veith, P. D. *et al.* Protein substrates of a novel secretion system are numerous in the Bacteroidetes phylum and have in common a cleavable C-terminal secretion signal, extensive post-translational modification, and cell-surface attachment. *J. Proteome Res.* **12**, 4449-4461 (2013).
144. Naito, M. *et al.* Determination of the genome sequence of *Porphyromonas gingivalis* strain ATCC 33277 and genomic comparison with strain W83 revealed extensive genome rearrangements in P. gingivalis. *DNA Res.* **15**, 215-225 (2008).
145. Cock, P. J., Chilton, J. M., Gruning, B., Johnson, J. E. & Soranzo, N. NCBI BLAST+ integrated into Galaxy. *Gigascience* **4**, 39-015-0080-7. eCollection 2015 (2015).
146. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421-2105-10-421 (2009).
147. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
148. Laine, M. L., Appelmelk, B. J. & van Winkelhoff, A. J. Prevalence and distribution of six capsular serotypes of *Porphyromonas gingivalis* in periodontitis patients. *J. Dent. Res.* **76**, 1840-1844 (1997).
149. Brunner, J. *et al.* The capsule of *Porphyromonas gingivalis* reduces the immune response of human gingival fibroblasts. *BMC Microbiol.* **10**, 5-2180-10-5 (2010).
150. Lamont, R. J. & Jenkinson, H. F. Life below the gum line: pathogenic mechanisms of *Porphyromonas gingivalis*. *Microbiol. Mol. Biol. Rev.* **62**, 1244-1263 (1998).
151. Honda, K. *Porphyromonas gingivalis* sinks teeth into the oral microbiota and periodontal disease. *Cell. Host Microbe* **10**, 423-425 (2011).

152. Hajishengallis, G., Darveau, R. P. & Curtis, M. A. The keystone-pathogen hypothesis. *Nat. Rev. Microbiol.* **10**, 717-725 (2012).